Strengthening the monitoring and evaluation of COVID-19 prevention programmes.

About this brief

This brief summarises lessons learned about the monitoring and evaluation (M&E) of COVID-19 prevention programmes. In this brief, we describe how M&E approaches have changed during the pandemic and outline 6 ways to strengthen future programmatic learning. The lessons shared in this brief are drawn from the work of the COVID-19 Hygiene Hub. These insights emerged from:

- Hundreds of informal conversations with programme implementers across 65 countries between April 2020 and May 2021.
- More than 50 in-depth technical support initiatives.
- More than 70 interviews with COVID-19 response organisations, donors and coordination mechanisms.
- 10 discussions with humanitarian organisations about common M&E challenges as part of a collaboration with the Global WASH Cluster.

This learning brief is designed to complement the COVID-19 Hygiene Hub resources on M&E and our list of external resources on this topic. This brief is primarily designed for people working within the water, sanitation and hygiene (WASH) sector and others who have been involved in COVID-19 prevention programming.

Patterns in M&E during the pandemic

In many countries, M&E was described as a 'casualty' of the pandemic. The table below summarises the challenges in programmatic data collection and learning, and how these changed as the pandemic progressed.

	Acute phase	Protracted phase
M&E challenges that arose during the COVID-19 pandemic	Urgent pandemic response action was prioritised over M&E (e.g. only implementation staff were given access to project sites). Difficulties adapting existing M&E frameworks to COVID-19 response due to new types of behaviours, the new settings in which they took place, the new delivery channels that were being used, and nature of COVID-19 transmission. A belief among programme implementers that COVID-19 was likely to only require a short-term response, therefore M&E was not considered a priority. Lack of guidance about the safety measures to consider during data collection. Concern about whether future data collection would be able to go ahead if restrictions were imposed. Remote data collection was hampered by unfamiliarity with possible methods, limited access to appropriate technology, and the absence of databases of target population contact details (e.g. phone numbers). M&E focused on documenting activities and reach rather than measuring acceptability, outputs, or outcomes. Data related to COVID-19 became politicized and this curtailed what information could be collected and how it could be shared and utilised. Operational research projects struggled to obtain approvals from ethical boards in a timely manner.	With multiple waves of the pandemic, data collection approaches were often delayed or had to be repeatedly adapted. Decision-making around the resumption of in-person data collection wasn't always systematic or based on epidemiological data. The design of survey tools didn't always lead to data which could usefully inform programming. Programmatic impact was hard to measure due to the high exposure to COVID-19 programmes and reliance on self-reported measures of behaviour and ad-hoc anecdotal programmatic feedback. Most programmes also lacked baseline data on behaviours, making it hard to draw conclusions about changes that may have happened. Fatigue with M&E approaches started to emerge among populations, causing reduced willingness to participate. In some areas data collectors were treated with suspicion, with populations concerned that they may contribute to the spread of COVID-19 or that divulging information to M&E teams may cause them to have to go into isolation.

"We thought it was better to do no M&E rather than take risks."

– Humanitarian NGO

"We have data on reach but no data and no evaluation of impact."

– Coordination mechanism

"We were not set up to go to remote M&E overnight. It took us months to build the system, train, and implement."

– Humanitarian NGO

"The situation was very unpredictable; we didn't know how or what to track. Something that was important today was no longer a priority tomorrow."

– Humanitarian NGO

"We went completely into reactive mode. I don't think anyone on the team was even thinking of monitoring."

– International NGO

"Safety concerns and precautions made everything but essential activities impossible. M&E was one of the first things to be taken out of the program."

– International NGO

M&E approaches were also affected by the broader challenges response staff faced when adapting to the 'new normal' way of working. These included changes to staff dynamics and changes in roles and responsibilities associated with remote working. Remote work tended to increase pressure on frontline staff who struggled to meet the requirements of responding rapidly, while also connecting to online meetings and dealing with the personal impacts of the pandemic on their lives. **Future response initiatives could benefit from re-evaluating how organisations support frontline staff and provide 'duty of care'**.

Organisations with poor M&E systems or lower capacity prior to COVID-19 were understandably at a greater disadvantage when the pandemic hit. This speaks to the need to **invest in M&E capacity strengthening initiatives to support routine monitoring and build resilience against future outbreaks.**

Areas of learning for strengthening M&E

The subsequent sections of this brief focus on six key areas of learning and identify common challenges and examples of positive practice. This brief focuses on the following topics: safe in-person data collection; effective and creative approaches to remote data collection; strengthening survey design; adapting processes for observing preventative behaviours; increasing routine operational learning; and identifying feasible alternatives to measuring programmatic impact. These topics have been selected not because they are the only considerations for doing M&E during outbreaks, but because they are the main areas of learning that emerged from discussions with practitioners.

1	**Doing in-person data collection safely**

While remote data collection became more commonplace during the pandemic, most M&E work still utilised in-person data collection approaches. Decisions around the use of 'standard', in-person M&E approaches were typically based on habits and ease rather than detailed risk assessments. As time went on many organisations, such as Action Contre la Faim, Solidarites International and Oxfam, developed phased programmatic and M&E plans where they forecasted how their interactions with communities would change according to government guidelines and community transmission. Other organisations reported reverting back to face-to-face data collection because they found that it was hard to build rapport with participants remotely or to create virtual spaces where opinions could be debated, or solutions could be brainstormed.

A degree of consensus emerged about the considerations for safe in-person data collection during the pandemic. These are summarised below.

General recommendations

☑ Align safe data collection procedures with national government guidelines and regulations.

☑ Inform participants about the data collection process including the safety measures you will be adopting.

☑ Avoid going inside people's homes and instead find a private outdoor or well-ventilated area.

☑ Keep interactions with participants as short as possible.

☑ Provide staff and participants with hand sanitizer during data collection.

☑ Ensure that staff and participants wear face masks throughout the data collection (provide these to participants if necessary).

☑ Make sure there is room for participants to maintain physical distancing throughout the session and cue this behaviour with the positioning of chairs.

☑ Reduce the number of shared objects (e.g. pens) and clean frequently touched surfaces.

☑ Train local data collectors from the target community to reduce the need for NGO staff to travel.

☑ Encourage frontline staff to get vaccinated as soon as possible.

Additional recommendations for Focus Group Discussions (FGD)

☑ Keep FGDs small, with no more than 6 people.

☑ Bring neighbouring people together to reduce mixing.

☑ Hold FGDs near people's homes to minimise the need for travel.

☑ Set up a handwashing facility in the FGD space or have hand sanitizer available so that participants and staff can clean their hands prior to and after the session.

☑ Avoid involving people who are over the age of 60 or who have pre-existing conditions. Instead use one-to-one methods or remote data collection to involve these people.

Image left: Oxfam staff working in the south of the Philippines adopted safety measures when undertaking interviews about the determinants of handwashing behaviour. This included provision of masks and physical distancing.

Image above: Data collectors from Y-PEER Sudan created safe and private outdoor spaces for Focus Group Discussions and put measures in place to minimise the risk of COVID-19 transmission.

2 Effective and creative approaches to remote data collection

Many organisations were unprepared for the sudden shift towards remote data collection, not just in terms of the technology it required, but also how methods would need to be adapted for new modalities. Adapting to remote data collection methods was reported to be easier for some humanitarian actors who had existing ways of reaching populations remotely due to pre-existing access limitations associated with crises. For example, in Nigeria the International Rescue Committee was able to utilise their phone hotlines and their existing contact networks with stakeholders in the community to understand the changing situation. Many actors mentioned they had prior outbreak response experience (e.g. in cholera or Ebola outbreaks) but that the learning related to M&E wasn't always transferable to the current pandemic, since movement restrictions were not common

during prior outbreaks and therefore in-person data collection was still the primary way of working.

The move towards remote data collection created challenges for obtaining quality data which could inform programming. Many organisations reported that their monitoring processes had been simplified due to the pandemic. Typically organisations focused on collecting numerical indicators to summarise programmatic reach, and self-reported perceptions and behaviours. Organisations felt the need to keep surveys short to maintain attention during phone or SMS-based surveys, however sometimes these reductions led to data that was not nuanced enough to act upon. Early on in the pandemic, many organisations tried using Interactive Voice Response Surveys (with pre-recorded automated questions and multiple-choice answers provided via the keypad) as this approach was able to be contracted out to service providers and allowed for large scale data collection with limited human resources. However, organisations often experienced high rates of no-response and the short, closed answer responses left response actors with more questions than answers when it came to making programmatic decisions. Other actors utilised social media to conduct polls or to promote surveys. This proved to be an efficient method of collecting large amounts of data, however this approach was often biased in terms of who participated, and it was difficult to draw any links between the population responding to the survey and the population living in the targeted areas.

Given that the information and responses to the pandemic were constantly changing, many organisations tried to establish monitoring mechanisms that would allow for data collection and sharing over time. Some organisations did this by embedding monitoring approaches in their remote communication platforms. One example of this was the 'action tracker' which was built into the U Afya, a mobile-based platform to build knowledge and motivation around COVID-19 prevention among mothers as agents of behaviour change in Kenya. Others placed an increased focus on monitoring perceptions related to COVID-19 and tracking how these changed over time. Organisations reported that qualitative data was often more useful to allow for rapid programme adaptation and identified many creative approaches to doing this remotely. Three novel examples of remote qualitative methods are provided below.

In India, Gram Vaani invited community members to use their mobiles to record voice messages about their pandemic experiences. These stories allowed them to adapt programmes and advocate on behalf of populations in real time.

In Honduras, GOAL organised video-based remote focus group discussions. In advance they provided participants with data credit and then invited older people and people with disabilities to 'buddy-up' with younger family members so that they could be part of discussions.

In Malaysia, ANAK wanted to engage and learn from hard-to-reach undocumented migrants. They found that WhatsApp was the best way of reaching them. They asked participants to do a 'photo journal of their daily life'. This involved participants sharing images of what they were doing to prevent COVID-19 with a short narrative description. This photo shows an outdoor sink built by a participant during the pandemic to facilitate hygiene upon entry to the house.

For those unable to do in-person data collection, phone-based interviews or surveys with populations were common. Below are some identified areas to strengthen phone-based data collection:

Tips for phone-based interviews or surveys

Anticipate issues with network coverage, power outages, phone charging and phone credit - This may include extending data collection periods, providing phone credit to participants, calling people on different days and at different times and notifying people by SMS in advance of your call. The Sudan COVID-19 Research Group together with Y-PEER data collection staff found that they had to be both patient and persistent to overcome the challenges of phone interviews in Sudan. They worked through their existing networks within communities to reach out to potential participants and let people know they were trying to get in touch. They adapted work schedules so that interviews could be done at times that suited participants (such as later in the evenings), and switched to using interview platforms that worked better with poor internet connection. In Tanzania, the NGO Maji Safi started by sending text messages to a large number of potential participants to ask preliminary questions. If a person responded several times, they then called them to engage in a longer phone survey. While this could introduce bias, it meant that those they called were willing and able to take part.

Use different types of questions – Typically phone-based data collection needs to be shorter than face-to-face methods as it is hard to retain attention otherwise. Mixing up different types of questions (e.g. comparative questions, scenario-based questions, normative questions) and responses (e.g. open answer, multiple choice or scaled or ranked response) can be a good way of maintaining attention throughout. Repeated interviews with the same participants can provide an opportunity for staggering a lengthy questionnaire over a set of calls.

Consider phone access and ownership – In many LMIC settings, phones may be shared among family members and phone ownership is often more common among men than women and less common among poorer individuals, older people and people with disabilities. This can affect not only who participates in phone-based data collection, but also who may overhear responses or influence a person's answers. Explaining the rationale behind data collection and confirming call times can help to overcome this. The Global Research and Data Support (GRDS) team at Innovations for Poverty Action (IPA) recommends comparing phone-based survey demographics to the demographics of prior surveys to understand how this might bias the data.

Focus on building rapport with participants – Without face-to-face interactions it can be challenging to connect with participants. In Lebanon research staff based at Oxfam found that they were able to build rapport during phone-based interviews by taking time to introduce themselves and the project thoroughly; matching female data collectors with female participants and vice versa for males; allowing time to listen to the concerns of participants (even if these were off topic); and by conducting repeat interviews with the same group of participants.

A Research Assistant at Oxfam conducts remote phone interviews with Syrian Refugees in Lebanon.

Data collectors also have a responsibility to inform – Given the unprecedented nature of COVID-19 and the associated changing guidelines, data collectors can play an important role in providing feedback to participants. In Zimbabwe, research staff based at Action Contre la Faim developed a set of messages about COVID-19 to share with participants at the end of phone-based interviews. These were shared when participants expressed views which were inconsistent with local guidelines or to inform them about available services. Sharing these at the end of the interview avoided adding a bias to participant responses to the interview questions.

Despite challenges with phone-based data collection, there were also advantages too. For example, it often made data collection more efficient as more interviews could be done per day across diverse geographical settings. For organisations who tried remote data collection for the first time during the pandemic, many felt that it was a positive shift and wanted to continue to strengthen capacities on remote data collection as part of their longer-term M&E strategies.

3   Strengthening survey design

Surveys remained the dominant mode of collecting data during the pandemic. However, it was relatively common for response actors to be unsure about how to adapt their programmes based on the data generated. Below we identify a few things that can be done to strengthen survey design and to facilitate useful programming insights:

Tips for strengthening survey design

Use validated and reliable indicators – Where possible, review surveys that have been developed and tested by others and utilise similar indicators or questions. Even though COVID-19 is a novel disease, there are standard indicators for some of the key prevention behaviours (e.g. handwashing) and tools developed during previous outbreaks that can be easily adapted (e.g. measures of perceived risk from prior Ebola or SARS outbreaks). Using standardised approaches can increase the validity and reliability of questions and allow for comparability of results. There are also opportunities to standardise indicators nationally and globally. For example, the National WASH Cluster in Colombia standardised core indicators and reporting among partners. This made it easier for all partners to share data from their departments on a monthly basis and for the cluster to analyze data to inform decision-making processes. The Global WASH Cluster has developed a WASH Sectoral Guidance on Covid-19 for Humanitarian Needs Overview that supports the development of such core sectoral indicators. The RCCE Collaborative Service is also working with a range of partners to standardise indicators around risk communication and community engagement.

Define the purpose of each question and how you will use the data – For each survey question, it is useful to define what it is measuring and ask yourself 'why does this question matter?' and 'how will I use this information to improve my programme?'. Doing this will allow you to prioritise the questions that are most likely to be of use for programmatic adaptation. In Syria, Save the Children were developing a survey for use with children in schools and temporary learning centres. They wanted the survey to be transferable and to respond to a range of different local contexts. They developed a list of potential questions and for each they listed the question type, the behavioural focus, whether it measured knowledge, attitudes or practices and the source of the question. This process allowed staff in each region to select questions most relevant to their programming.

Include questions that are designed to measure change – To facilitate the adoption of COVID-19 prevention behaviours, it is important to understand what has changed in people's circumstances (e.g. behavioural determinants) or how actual behaviours have changed over a defined time. Therefore it is important to design questions so that they focus on this aspect of change. For example, WaterAid conducted a multi-country rapid assessment of hygiene behaviours in 8 countries. This included indicators related to exposure to programming, preferred delivery channels, self-reported behaviours and determinants of these behaviours. WaterAid realised that to measure change, it was important to be precise during data collection. They found it useful to differentiate between 'normal' critical moments for hand hygiene and 'new moments' that were promoted during the pandemic (e.g. Handwashing with soap before entering or leaving the household, after coughing/sneezing, after touching frequently touched surfaces, and before/after caring for someone with COVID-19 symptoms). This allowed them to understand how COVID-19 prevention activities could be integrated into their existing hygiene programming and how to adapt programmes as the pandemic continued. Asking people about when they changed their behaviours and why can also be key to making programmatic decisions.

Validate the understanding of questions in local languages – Many aspects of epidemiology and disease perception and prevention are complex to explain. Before rolling out a survey, it is important to dedicate time to getting the local translations of terms right. Piloting the survey with a few individuals often allows you to pick up when questions are being misunderstood. For example, an NGO in Uganda initially developed their survey tool in English, but then moved to translate it into two local languages to improve understanding. They used an approach of working with native language speakers to brainstorm appropriate local terms for some of the epidemiological or COVID-19 specific concepts. Cognitive interviewing can be an easy method to check whether respondents understand the questions as intended by the investigator. To apply this method, select the key word(s) from each question and request the respondent to explain that word to you. If the explanation that

the respondent gives matches the meaning which the investigator intended, this suggests that the key word is well understood. For testing entire questions, the respondent can also be requested to answer the question and then explain their thoughts behind their response. In this method, misunderstandings are effectively identified.

Collect socio-demographic data – Pressure to reduce the length of surveys can often result in actors cutting questions related to socio-demographic factors. However, information about gender, age, location, abilities, education, economic status and other factors can often be key to translating insights into targeted programmatic actions. Questions related to economic status and religious background can be sensitive to some people or in some countries. Using previously validated questions, piloting surveys and taking time to build rapport can mitigate these issues. Collecting and analysing socio-demographic data can also identify equity barriers within data collection and future programming (e.g. fewer female participants, limited representation from people with disabilities, etc). Overcoming these barriers to data collection may require stratified or purposive sampling and consideration of phone or social media access.

Include some open-ended questions and complement surveys with other tools – One of the challenges with surveys is that they can only generate data on the specific questions asked about and normally users are required to specify potential answer options in advance. This means that findings can sometimes overlook other challenges that were not asked about. Including some broader open answer questions can help to overcome this and generate information about why people think or behave in the way they do. However, open-ended questions need to be carefully selected and prioritised as each response will require a more time-consuming analysis. Including open-ended questions among a smaller sub-sample of participants may make this process more feasible. Alternatively, the findings from survey data can be complemented with other data collection methods. For example, survey data might be usefully followed by a short period of qualitative data collection to explore and validate some of the patterns identified. Using multiple methods can also address the limitations and biases of each method.

Make a data analysis plan – Often data analysis comes as an afterthought. A data analysis plan should outline exactly what is done with the data from each question and should consider timelines for the analysis and staff capacity. If you plan to look at the combined effect of several variables on a particular outcome, this needs to be planned from the outset. If survey tools are standardized across countries, analysis plans can also be standardised allowing for greater efficiency. Similarly it is important to plan in advance for how findings may be disseminated and how insights will be used to inform programming. This may include informing other stakeholders early on about the data collection you have planned and allowing sufficient time and budget within programmes to make iterative changes.

4 Adapting processes for observing preventative behaviours

Observation is generally considered to be a more reliable measure for understanding behaviour than self-report. This is because with frequent messaging about COVID-19 preventative behaviours, people often over-report their preventative actions because they want to be seen as someone who does the 'right' thing. In the early phase of the pandemic, observation was often avoided due to concerns about safety. However, over time we have seen effective and safe adaptations of observational methods. Common adaptations included adjusting methods to suit observation in public settings, allowing for observational data to be collected by community members, and conducting observations over a short duration of time. An increasing number of organisations have also been using spot-checks or observational checklists. These provide rapid assessments of the physical environment to indicate whether it is conducive to the practice of prevention behaviours. For example in Indonesia and Mozambique, SNV worked with Upward Spiral to develop a checklist that could be used to assess COVID-19 prevention measures in marketplaces, transport hubs and health care facilities. The checklist allowed managers to actively engage in prevention of COVID-19 in these spaces by allowing them to calculate a COVID-19 safety score and receive recommendations about specific things they could do to improve their score. Rewards were provided to incentivise changes. Such checklists are relatively rapid to conduct, allowing the same information to be monitored over time. For example, WaterAid developed a standardised checklist to monitor the functionality and accessibility of handwashing facilities that they installed in public places in 8 countries and intend to continue to use this repeatedly over time. Below we identify a few things that can be done to strengthen observational methods.
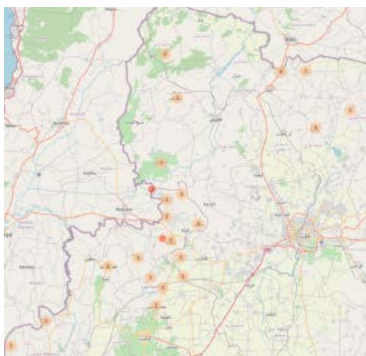
Image left:
A map of the public locations (marked in orange) in North West Syria where UDER conducted observations of mask use behaviour.

Image right:
An observer documents mask use as people enter a marketplace.

Tips for strengthening observational methods

Invest time in building staff capacity – Observation is a new skill for many and requires both classroom-based training and applied practice in real world settings. In Ethiopia, the Democratic Republic of the Congo and Bangladesh, Oxfam trained their staff on both observation and spot-checks to monitor the use of their Oxfam Handwashing Stand in displacement camps. They found it was useful to create

mock scenarios for classroom training and then give staff in each country time to practice the tools before sharing feedback and reflections. The quality of observations can also be strengthened by getting supervisors to conduct random checks of those conducting the observations and by facilitating team reviews of the data collected each day. This process can also help to overcome any contextual barriers that may have not been anticipated.

Provide guidance on how to classify observations – Observers are typically asked to categorise whether a person practices a behaviour at a particular moment and potentially some details on how they do it. To facilitate this classification it's important that categories are well defined. For example, in North West Syria UDER conducted a survey on mask use and then complemented this with observations in public settings. To aid with classification, they included images when training staff to explain what 'correct' mask use looked like and various forms of 'incorrect mask use'. Following piloting in the local context an additional category was added for niqab/shemagh use (cultural face coverings) as for people wearing a niqab/shemagh it was not normally possible to tell if they were wearing a mask underneath.

Capture the right amount of detail – As with all data collection methods, it's important to only collect data that will be useful for programming. This requires limiting data collection to key variables and considering the level of detail that will be necessary for decision making.
Kenya NBCC provided over 5,000 handwashing facilities in public settings throughout Kenya. Their evaluation included checking the presence of water and soap at facilities. Initially they were recording detailed measures of how much water was in the handwashing units (¼, ½ full etc.), but in hindsight, they realised they could simplify this to record whether water was present or absent. They realised that this binary categorisation was enough to indicate whether the facility was functional or not.

Spot Checks were routinely conducted at the handwashing facilities distributed by NBCC to assess maintenance and functionality of the stations over time.

Agree on a way of measuring the denominator for your outcome – Observational data can be used in a range of ways, but commonly it is used to calculate the proportion of people practicing a behaviour at a key moment or in a particular setting. At the beginning of data collection, it's important to identify how the denominator will be calculated and pre-tested. For public settings this can be challenging and may affect which sites are selected for observation. In Indonesia, UNICEF and the National Government developed a real-time observation-based monitoring system to assess handwashing, physical distancing and mask use. They used a network of volunteers to conduct rapid observations across the country and enter data into a standardised template. UNICEF decided to focus their observations on settings where there was a clear entrance point, as this allowed them to capture the number of people entering the space. They trained their volunteers, many of whom were part of the National COVID-19 Taskforce, to document the prevention behaviours of the first 10 people they saw entering that space. This made the approach rapid, feasible and easy to measure. Volunteers were also provided phone credit to support and incentivise their work.

5 Increasing routine operational learning

Many response actors mentioned that in addition to the use of formal M&E methods, they found that informal information sharing helped to rapidly adapt their programmes. Informal operational learning was particularly key in the early stages of the pandemic where large-data collection was less feasible, and the situation was changing rapidly. Designing programmatic reports or regular meetings in a way that can facilitate programmatic learning is an approach that will be key in establishing stronger M&E systems for the future. Below are four simple ideas to strengthen operational learning:

Tips for strengthening operational learning

Use all available existing data – In the early phase of the pandemic many actors overlooked the potential to utilise existing data to inform the first phase of their programming. This could have included the utilisation of prior assessments of relevant behaviours (e.g. handwashing), experiences related to other disease outbreaks (e.g. cholera or Ebola) or the availability of relevant infrastructure and services (e.g. water services). In Zambia, GRID3 worked with the National Government to create the Zambia Data Hub which used existing data (e.g. Demographic Health Surveys and government data) and made it accessible through online mapping applications and dashboards which allowed users to visualize populations at risk (such as those with limited access to water for handwashing, limited health care access, and areas with high population density).

As time went on, the Zambia Data Hub expanded to capture new data and allowed for visualisations of perceptions and behaviours based on surveys submitted from the organizations conducting COVID-19 community engagement campaigns. It also included COVID-19 cases at district and provincial levels, locations of testing sites and vaccination centers and tracking of vaccine doses administered. This centralised portal for knowledge sharing allowed local actors to coordinate timely action at each stage of the response. Humanitarian WASH Clusters have also developed guidance on how to conduct secondary data reviews to support partners who want to utilise existing databases to inform their COVID-19 response work.

Meet frequently with frontline staff to reflect on programming and capture this in reporting – Under normal circumstances, programmes are sometimes changed as a result of informal discussions that take place on a day-to-day basis within organisational offices. With many staff working remotely, it was easy for informal learning opportunities to be missed. Oxfam, Action Contre la Faim and partners have been using the Community Perception Tracker approach in 14 countries. This encourages staff to employ active listening skills and systematically document perceptions shared with them during their ongoing programming activities. They realised that the process was substantially strengthened by having weekly meetings between programme staff. This allowed them to translate insights emerging from communities into agreed ways to improve programming. They also moved towards standard reporting templates that guided staff on how to summarise information in ways that would resonate with other actors and maximise learning.

Create open dialogues with stakeholders – Many actors have indicated that the most valuable mechanisms for ongoing programmatic learning have been through informal networks that have been set up with key community stakeholders. This can facilitate two-way information sharing and increase the acceptability of other data collection approaches. For example, in Kenya the International Rescue Committee initially faced reluctance from communities to participate in data collection within displacement settings. They worked with other response partners to hold frequent joint meetings with community leaders to generate a greater understanding of the rationale behind data collection. To make monitoring less burdensome on the community, the partners tried to harmonise or use joint data collection procedures where possible. Once these communication forums had been established, acceptance of M&E increased among the community and community stakeholders were able to share feedback upwards to improve programming. Since in-person data collection has remained challenging throughout the pandemic, many organisations have also taken time to train community members on data collection. This has proved particularly valuable for monitoring functionality of handwashing facilities and ensuring water and soap are present as these same community stakeholders can often take direct action to address challenges.

Share results creatively – The sharing of operational learning within and between organisations needs to be timely so that it can influence programming, but it also needs to be well described and formatted. Achieving these things is hard to do in a fast-paced crisis. One common challenge with dissemination is that data collectors don't clearly describe how data was collected and what the intention behind the methodology was. Without this information, results can often be misinterpreted. When developing dissemination documents, organisations sometimes struggle to draw connections between patterns emerging from their data and their recommended set of actions. Lastly, findings are often lengthy or presented in a way that it is hard for potential users to digest or make sense of. In Cox's Bazar in Bangladesh, there has been an ongoing effort to develop rapid briefs on qualitative data collection from populations. To make findings more engaging, the partners worked with a graphic artist to illustrate some of the experiences they learned about. Since the final product was more visually engaging than the standard types of information being shared, it helped the findings to get more attention among stakeholders who could act on the findings. Using existing networks such as RCCE coordination groups or WASH Clusters can be an easy way to share insights with those who can utilise the results.

6 Identifying feasible alternatives to measuring impact

Many COVID-19 response actors faced challenges in evaluating the effectiveness of their programming. There are several factors that made measuring impact more complex during the pandemic. These include:

Multiple factors affecting COVID-19 transmission – Some actors considered monitoring COVID-19 cases, positivity rates or mortality as a potential outcome measure. However, there are a range of limitations with these measures and multiple factors that may influence transmission at any time point. Therefore, this wasn't considered a viable indicator of programmatic impact. However, in some countries, where COVID-19 testing was limited, the monitoring work of NGOs and of communities themselves have been a useful source of information about real-time transmission and mortality and allowed response actors to adjust programmes accordingly.

Scale and diversity of the response – Sometimes aid and development actors are working in a setting where there are few other interventions which are trying to achieve the same public health outcome. This makes it easier to attribute changes in behaviour or perceptions to the programme. However, the COVID-19 pandemic triggered responses at an unprecedented scale and across all settings populations were exposed to the programming of multiple government, non-government and community-led initiatives. Therefore, any changes observed are likely to be attributable to the combined impact of all these response programmes, rather than one.

Increased use of remote delivery channels - Many actors were using delivery channels that they were less familiar with, including social, digital and mass media. While some actors found ways to monitor the reach of their messages (e.g. through social media analytics or media monitoring) it was much more challenging to gauge the impact of messages on people's thinking and behaviour.

Reduced frequency of data collection and reduced variety in the types of data collection methods – Many organisations reported relying on self-reported measures of behaviour. While they recognised the limitations of these approaches, they considered alternative methods to be infeasible or unsafe in their contexts. Others said that the reduced interactions with populations meant that data collection only happened at a few select time points and this made it hard to understand how behaviours changed over time. This was considered particularly important during the pandemic given that behaviour appeared to be changing regularly in response to changing evidence and guidelines.

A World Vision field worker conducting a quantitative interview face-to-face in a refugee camp in Zimbabwe

Tips for strengthening COVID-19 programme evaluation

Below are some ideas to help strengthen COVID-19 programme evaluation:

Use a theory of change and develop indicators to track each stage of the hypothesised mechanisms of change - A theory of change describes how a project proposes to bring about a change in behaviour or health outcomes by outlining a step-by-step series of causal events. Theories of change are normally developed as part of programme design processes but are valuable to inform monitoring too. For example, in the Tongogara Refugee camp in Zimbabwe, RANAS worked with UNHCR, World Vision and SDC to design their programmatic monitoring to reflect each level of their theory of change. To assess their intervention delivery, they measured the recall of interventions among the population. To assess outputs and outcomes, they conducted a survey to monitor changes in behavioural determinants and self-reported handwashing practices and physical distancing practices. By combining these indicators they were able to understand whether their programme was implemented as intended and had the intended effect.

Use more than one approach to measure behaviour - Given that access to communities can change frequently over the pandemic, many actors have found that it is helpful to have multiple indicators or methods to measure the same behaviour. In India, the Janseva Gramin Vikas Va Shikshan Foundation and Ranas used a mix of questions to understand mask use behaviour. They started off with an open-ended question: "Imagine you are leaving the house to go shopping or going to visit somebody. What do you do?" As participants answered the enumerator listened and ticked whether putting on the mask was mentioned. Later in the questionnaire they asked:"In which situations do you wear a face mask?" Data collectors probed respondents based on specific times such as leaving the house or catching public transport. Combining both questions provided a more reliable measure of behaviour and allowed for an understanding of behaviour as practiced within daily routines. It can also be useful to ask participants normative questions about their views on the behaviour of others in their community.

Measure the acceptability, relevance and sustainability of programmes - These factors are often overlooked in programmatic M&E approaches and are comparatively easy to measure. Many actors mentioned that in hindsight they wished they had included more qualitative indicators to understand 'why' and 'how' their programme was realising an impact.

Joint monitoring initiatives – Rather than looking at the impact of one programme or organisation, it makes sense to pool resources and look at the collective impact of response initiatives during the pandemic. In many regions of the world, existing coordination structures have helped facilitate joint monitoring initiatives. For example, the National WASH Cluster in Palestine developed a Vulnerability Ranking System to help partners identify needs and particularly vulnerable regions. They adapted this to capture COVID-19 data so that they were able to track how the pandemic affected the vulnerability of different regions. In other regions of the world, such as Indonesia, Government and NGO partners with limited time and capacity reached out to research institutes to facilitate quarterly rounds of formative research on COVID-19 prevention behaviours. Using an external neutral partner facilitated trust in the findings which were then used by all response actors.
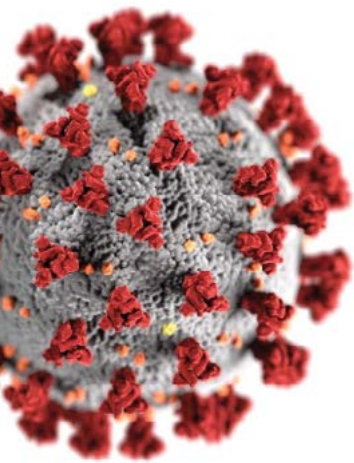
COVID-19
HygieneHub

hygienehub.info

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE   CAWST   Wash'Em